

Міністерство освіти і науки України Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» Фізико-технічний інститут

«Криптографія» комп'ютерний практикум №1

Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконали:

студенти групи ФБ-93 Приходько Андрій Шахова Катерина

Мета роботи:

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела

Завдання:

- **1.** Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення (10) H, (20) H, (30) H
- **3.** Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

Джерело	Н	R
Н1 текст з пробілами	4,378	0,124
Н1 текст без пробілів	4.45	0,11
Н2 перехресна біграма з	3,97	0,206
пробілами		
Н2 не перехресна біграма з	3,9766	0,204
пробілами		
Н2 перехресна біграма без	4.1464	0,17
пробілів		
Н2 не перехресна біграма без	4.1453	0,17
пробілів		

Частоти літер у тексі з пробілами :

- {' ': 0.15881530158478815,
- 'o': 0.09154867626424572,
- 'a': 0.07024992412546349,
- 'e': 0.06565016780368509,
- 'и': 0.059094167997220136,
- 'н': 0.05403036740869764,
- $'\pi'$: 0.04894457420089818,
- 'p': 0.04708179934813876,
- 'T': 0.0462713713277824,
- 'c': 0.04283282530382804,
- 'в': 0.033971998997136564,
- 'м': 0.027240058236455846,
- 'к': 0.027008036103083805,
- 'д': 0.026505504752605026,
- 'y': 0.025483947587189738,
- 'π': 0.024256759431534776,
- 'r': 0.019380995737830385,
- 'ь': 0.017025586213266827,
- 'я': 0.016944213521942038,
- '3': 0.016293231991343703,
- 'ы': 0.013707999595335805,
- '6': 0.013139490387026114,
- 'ч': 0.012006870494262125,
- 'й': 0.007763394604770639,
- 'ж': 0.00741041306537526,
- 'ш': 0.007121210121883096,
- 'x': 0.006538405711043374,
- 'ю': 0.004483195439610467,
- 'щ': 0.0029668043404633406,
- 'u': 0.002624819110706447,
- 'э': 0.0018616752217955653,
- 'φ': 0.0017462139705914695}
- H1 = 4.378

Частоти літер у тексі без пробілів :

- {'o': 0.10881487854264225,
- 'a': 0.08351033761581432,
- 'e': 0.07805986770045001,
- 'и': 0.0702895211288935,
- 'H': 0.06422546537874564,
- $'\pi'$: 0.0581614096285978,
- 'p': 0.05597599965534753,
- 'T': 0.05500470633390297,
- 'c': 0.05091587477104763,
- 'в': 0.04039875246250273,
- 'м': 0.032388193564137344,
- 'k': 0.03210881618404442,
- 'д': 0.031499146854589294,
- 'y': 0.03027980819567905,
- $'\pi'$: 0.02884245074015423,
- 'r': 0.02302252382214366,
- 'ь': 0.02023266652654281,
- 'я': 0.020146503409317888,
- '3': 0.01937364635784587,
- 'ы': 0.016297884173271417,
- '6': 0.015620328751457266,
- 'u': 0.014269134413157369,
- 'й': 0.009232508560827897,
- 'ж': 0.008809525985360103,
- 'III': 0.008458346007579743,
- 'x': 0.0077677355680042404,
- 'ю': 0.005335585759064425,
- 'щ': 0.0035326878062217604,
- 'ц': 0.0031279822556198587,
- 'э': 0.0022193530194297827,
- 'φ': 0.002078358827607185}
- H1 = 4.45

Частоти перехресних біграм у тексті з пробілами

4 0.000 0.0000 0

Частоти не перехресних біграм у тексті з пробілами

Таблиця частот перехресних біграм без пробілів

Таблиця частот не перехресних біграм без пробілів

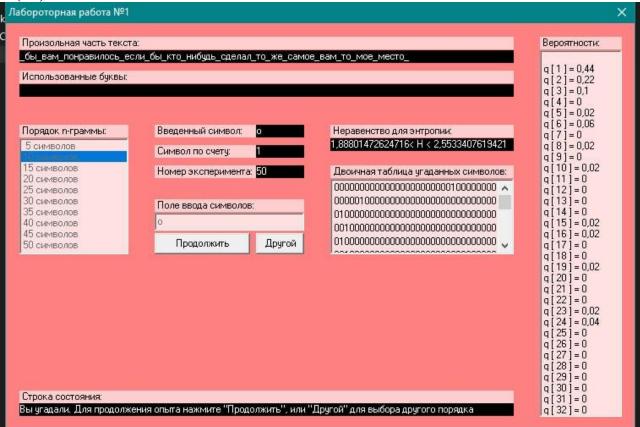
a 0,0003 0,0015 0,0047 0,0025 0,0035 0,0018 0,0011 0,0040 0,0040 0,0041 0,0007 0,0055 0,0118 0,0041 0,0097 0,0055 0,0118 0,0041 0,0097 0,0055 0,0118 0,0091 0,0047 0,0055 0,0118 0,0091 0,0047 0,0095 0,0097 0,0095 0,0097 0,0090 0,0097 0,0090 0,0097 0,0090 0,0097 0,0090 0,0097 0,0090 0,0097 0,0090 0,0097 0,0090 0,0097 0,0090 0,0090 0,0097 0,0090 0,0097 0,0090 0,0090 0,0097 0,0090

Таблиці найбільш частих біграм для різних джерел :

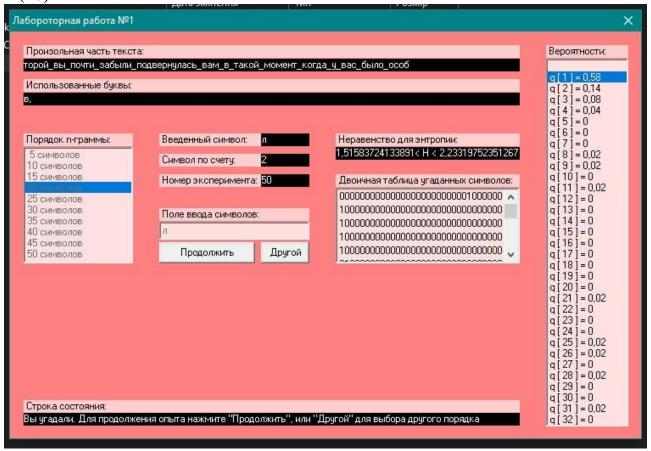
Перехресні з пробілами		не перехресні з пробілами		Перехресні без		Перехресні без пробілів	
				про	білів		
a""	0,0179	a""	0,0179	ал	0,0127	ал	0,013
e""	0,0154	e""	0,0153	на	0,0128	на	0,0127
o""	0,0188	o""	0,0186	он	0,0125	ОН	0,0124
""в	0,0141	""в	0,0142	то	0,0128	то	0,0129
""н	0,0148	""н	0,0149	но	0,0112	но	0,0113
""o	0,0117	""o	0,0117	по	0,0114	по	0,0113
""п	0,0168	""п	0,0166	не	0,0106	не	0.0105
""c	0,0139	""c	0,0138	ри	0,01	ри	0,01
д""	0,0154	д""	0,0154	др	0,0106	др	0,0107

Завдання 2

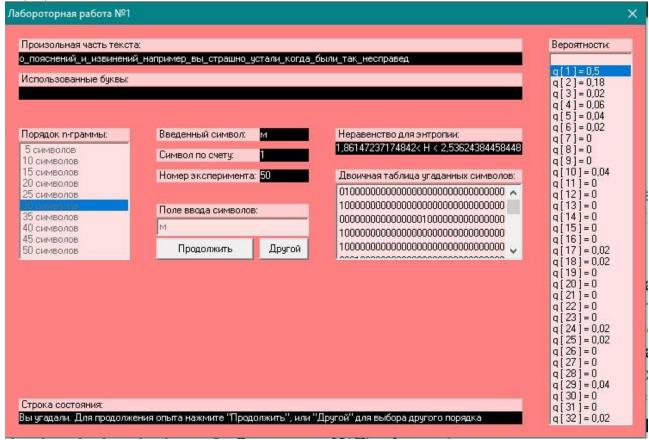
H(10)



H(20)



H(30)



Завдання 3.

Оцінка надлишковості мови:

H(10)	0,5494 < R < 0,6224
H(20)	0,5534 < R < 0,6968
H(30)	0,4928 < R < 0,6278

Висновок: під час виконання комп'ютерного практикуму ми засвоїли поняття ентропії на символ джерела, його надлишковість. Навчилися вимірювати на практиці частоти символів та біграм у тексті. Зрозуміли як визначати ентропію та надлишковість мови.